On the Syntax and Semantics of English Modals*

Shuan-fan Huang

*Sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

On the Syntax and Semantics of English Modals

1. In approaching the problem of the semantics of utterances in natural languages, it seems natural to assume assertions or statements as the basic class of messages and to derive other classes from them. We then compare various types of messages with kinds of expressions found in languages and take note of points where correlation is not symmetric. We could alternatively begin with various forms of expression and study the nature of messages they convey. Finally, we can, as do Austinian philosophers, consider the use of words and note the differences between what is presupposed by the use of these words and what they can be used to assert, to order, to promise, ato accuse and so forth. A number of philosophers have tried to analyze the so-called happiness condition for the performance of certain kinds of linguistic utterances. A grammarian's job should be to figure out how illocutionary forces and happiness conditions can be related to certain lexical and syntactic properties of sentences.

Within the tradition of Aristotelian logic, sentences are dichotomized into those to which there is truth value and those to which there is none. But the truth value test is far from unambiguous. No truth value is assignable, for example, to (1) or (2):

(1) He would have been killed

(2) I ought to have read that yesterday

All imperatives and interrogatives don't have truth value, though answers to some questions, e.g. (3) or (4)

(3) Who's the author of Tropic of Cancer?

(4) Are you ready yet?

will have. An answer like

(5) He might come tomorrow

represents a point on a scale of several possible answers to the single question

(6) Will he come tomorrow?

It seems clear that the truth value test cannot even begin to be used as a means to an interesting taxonomy of utterance types.

Strawson and Searle have tried to explicate Austin's notion of illocutionary force in terms of Grice's theory of meaning. On Grice's account, a speaker S means something by an utterance Y if and only if in uttering Y the speaker S intends to achieve some effect in some hearer H and that H recognize S's intention and that this recognition will function as H's reason in a certain intended manner. Austin claims that there are "third power of ten" illocutionary forces in English. This is important to his conception of illocutionary acts. Illocutionary forces may be more or less indeterminate. Suppose I ask you to do something for me. My utterance can be a request, an entreaty or a plea. One might think of illocutionary acts as on a continuum of specificity but this would not do justice of the full complexity of the speech acts. For under the rubric "illocutionary force" are all sorts of different principles of distinctions: purposes of acts, relations between speaker and hearer, degrees of commitment and roles of acts, etc. Consider for a moment the relation between subject-person and illocutionary force by looking at some simple past statements in the third person, second person and first person:

(7) John went to the hospital this morning

(8) You went to the hospital this morning

(9) I went to the hospital this morning

Clearly (7) is most likely an assertion, with speaker's knowledge based on either direct observation or on reliable evidence. The claim for reliability in (9) is much stronger and normally should be beyond any shadow of doubt. (8) is

not just an assertion; it is also likely to be an accusation. The speaker is challenging the hearer to the contrary. It obviously cannot have the force of (7) and (9), of teaching the listener something he did not know before.

The present tense involves the question of the shared knowledge of speaker and hearer. In

(10) I know that he left

"I know" signals the trustworthiness of a statement made in the best evidential conditions. It functions like adverbs or parentheticals, that is, as if it said

(11) He certainly left.

(10) therefore commits the speaker to the truth of the statement he left. It can be contradicted by

(10a) No, you don't, because he did not leave

(10b) He certainly left, but you did not know

(10c) You may have thought that he left, but you did not know

Performatives, however, cannot be contradicted without creating a bizarre communication. I consider (13) and (15) as pathological.

(12) I promise you to do it

(13) No, you don't

(14) I order you to go

(15) No, you don't.

2. Many sentences in the third person have quite different meanings from those in the first or the second, as indicated above. This is seen most clearly in modal sentences, with which this paper is centrally concerned. As discussed in logic, the notion of modality is first due to Aristotle, who argues for two basic modalities only--possibility and necessity (approximately may and must)--various others being

reducible to these two in one way or another. One thing that emerges clearly from Aristotle's discussion is that there are essential similarities among modal sentences and quantified sentences. If a sentence is necessary, it is true of all possibilities; if a sentence is possible, it is true of some possibilities; if it is impossible, it is true of no possibility (it is not true of any possibility). We may note further that the notion necessity may be related to obligation in the same sense as impossible is related to prohibition.

The English modals can and may, capable of meaning either possibility or permission, and must (and it negation), meaning either necessity or obligation (and impositibility and prohibition) point up exactly these parallelisms.

The following tripartition captures the above-mentioned similarities:

A. all-some-none (quantified mode)

B. necessity-possibility-impossibility (alethic mode)

C. command-permission-prohibition (denotic mode)

Categories in C may further be thought of as results of adding to categories in B an element of will with regard to another person, implying that at least two persons are involved in any sentences that belong in these categories.

Aristotle was concerned with other implications, however. He developed a theory of logical relations of sentences containing such modals, a topic which has been treated ever since, but does not concern us here.[1]

---

[1] Interested readers are referred for more information on this topic to (to mention but one) I. M. Bochenski, Ancient Formal Logic, North-Holland Publishing Company, Amsterdam, 1951.

The term modal is often applied to the closed set of auxiliaries in English consisting of can, could, may, might, must, will, would, shall, should, ought to and sometimes need (not) and dare (not). Other linguists also recognize have (got) to, be to, be able to, had better, had/would rather and one or two more.[2]

---

[2]Long (1961, p. 138) accepts have and be as true auxiliaries, but regards the modals as full verbs capable of taking sentential objects. Joos (1964) admits the modals partly on the basis of their behavior with respect to do and partly on semantic grounds. Diver (1964) includes keep and used to but excludes dare and need.

Each of Aristotle's two basic modalities can be enriched in a number of interesting ways to correspond to diverse needs of human communication. Possibility, for instance, can range all the way from mere possibility to near inescapability. Necessity may be attributed to such unrelated factors as laws of logical inference, to physical laws, to human will, or to moral obligation of all sorts. In fact, languages tend to treat logical necessity indistinguishably from physical consequences or moral obligation. Thus, in uttering

(16) John ought to do it

(17) You must open the window

it is more often than not really inescapable that John or you do it. Indeed, with ought, it is almost always the case that we fail to do it, and in present or past tense, ought in fact presupposes the falsity of predication.

(18) John ought to be here by now

(19) You ought to have been here this morning

mean that John isn't here now and that you failed to show up this morning.

Theoretically, modalities must be combinable, since we can say such things as

(20) It must be the case that he can do it

(21) It may turn out that John will have to go

In English, however, modals are mutually exclusive, at least those that are most readily accepted as true modals. Thus

(22) *must can

(23) *dare (not) will

(24) *must be to

(25) *ought to must

(26) *may must, etc.,

are never permitted.

Let me quickly name some of the major characteristics of English modals before going on to a somewhat more detailed examination of their syntax and semantics. The first characteristic of a modal is that under negation, the negative particle not follows the modal; in contrast, a non-modal verb, when negated, calls for do-support and the particle not is then attached to the auxiliary do. The following are impossible in English:

(27) *I like not John

(28) *we saw not him

Secondly, the inversion transformation obtains for modals under interrogation or after the negative preverbs such as scarcely, seldom, never, hardly, etc. Thus

(29) Will they be there?

(30) Ought we to ask them?

(31) Seldom can they see the light.

Third, modals can only occur initially in a verb phrase, a characteristic that is shared by no other verbs in English. Thus whereas

(32) I want to begin.

(33) I begin to want.

(34) I ought to begin.

are O.K.,

(35) *I begin to ought.

(36) *I want to ought.

are definitely out. This characteristic seems to be correlated with the total lack of selection restrictions on the part of English modals. For every sentence in the language, it is possible to create a modal sentence by the simple process of putting a modal, with appropriate sense inflections, before the main verb. From

(37) John is reading a book

comes

(38) John may be reading a book.

From

(39) The table is red

we get

(40) The table may be red.

The addition of modals in no way affects the grammaticality, nor the selectional restrictions of the original sentences, which are taken intact from the deep structure. Fourth, all modals, including such morphologically past tense forms as could, might, should, would may refer to the future and may co-occur with future time adverbials. There is, for instance, no time difference in the following:

(41) He may go tomorrow--He might go tomorrow

(42) I shall ask him--I should ask him

(43) Can you help?--Could you help?

In indirect discourse, only past tense forms are used, of course. But must, ought to, (and dare, need) don't change even in indirect discourse.

Fifth, sentences containing modals passivize across both infinitive and preceding verbals, which is not the case with other complement-taking verbs (with the exception of a small class of intransitive verbs; seem, happen, appear, etc.) like want, avoid, expect, endeavor, like, etc.

(44) John may see Mary

is passivized across to

(45) Mary may be seen by John

and the meaning is preserved. But

(46) John expects to see Mary

and its passive

(47) Mary expects to be seen by John

are completely different in meaning.

Returning now to how English implements modals to effect the idea of possibility, we note that among modals expressing possibility of various shades, can and may are most deserving of attention. Since could and might, morphologically their past tense forms, are chiefly used in a tentative sense to make less positive statements or more polite requests, and semantically are not too distinct from can and may, what I have to say below concerning the latter will also be applicable to the former, unless otherwise specified.

One sense of can is concerned with ability, of whatever type. In this sense, it is not used with future time adverbials to refer to the future; future time is indicated by will be able to; could refers to past time.

(48) When he is older, he can run a mile

is odd, but

(49) When he was young, he could run a mile

is well-formed.

Can also expresses feasibility or the absence of anything to prevent from occurring. It is replaceable by may and can refer to the future. But can in this sense is not replaceable by may in questions. (50) is not the same as (51).

(50) Can he be hiding?

(51) May he be hiding?

The past time analogue of feasibility can is can have, not could. Contrast (52) and (53)

(52) He can be hiding.

(53) He can have been hiding.

where the difference is only in time. In (54)

(54) He could be hiding

the idea of feasibility is much less positive. There is no difference in time.

In negative sentences and interrogatives, can appears where may would be likely or almost certain in the corresponding affirmative sentences:

(55) He can't have left: He may have left.

(56) These figures can't be right: These figures may be right.

(57) Who can that be?: That may be John.

In He can't have left, the idea of possibility is present in time (it isn't possible) and the idea of leaving is past (that he has left). With can, there is a contrast between (58) and (59).

(58) You can't go

(59) You can not go (do what you please: you can go or you not go)

Here can't negates the ability (or permission) to act; can not positively states ability (or permission) not to act.

May, like can, is a full predicate word expressing possibility of various types. Most often it expresses a kind of possibility that involves uncertainty on the part of the speaker, much as the adverb perhaps does. May is used with reference to both present and future; may have is the past time analogue; might is used in a tentative possibility sense.

May is also used to give permission; reference may be to the present or future time. There is no past time analogue (why?); might is available only as, again, the analogous tentative form in request-questions (might he go?). (60) in the permission sense is ruled out.

(60) *You might go

In interrogatives, may is confined to the permission sense. The question corresponding to (61) will employ some such locution as (62).

(61) He may (possibility) go home

(62) Is there a possibility of his going home?

The may which recognizes uncertainty is not negated. In (63)

(63) You may not like it

there is no negating of may--what is negated is the following infinitive, like. Permissive may can be negated, as in (64).

(64) Cars may not park here.

The two senses of may and the different scopes of not intersect to yield logically four possible interpretations to a simple sentence like (65).

(65) He may not read that.

It is, however, only two-way ambiguous between (66) and (67)

(66) He is not allowed to read that

(67) It is possible that he'll not read that.

Where defective may seems inadequate for the purpose at hand, other locutions can come to the rescue: there is a chance, it is possible, it is permitted, it is allowed, etc.

Sentences containing stative verbs or adjectivals like (68)

(68) John may (permission) be tall

may seem to be odd. Under different circumstances, it would be perfectly natural. The sentence

(69) They agreed that in the play John might be tall but Mary had to be short.

is impeccable. Similarly, sentences containing non-human subject and adjectivals like (70)

(70) The answers may (permission) be correct.

may be rejected at first glance. Embedding it to another sentence, we obtain a well-formed sentence.

(71) It would be incredible for a teacher to tell students that the answers may be correct

or may be wrong.[3]

---

[3]Since it may be the semantic content of a sentence embedded at n^{th} depth or conjoined at n^{th} branching which determines the ultimate acceptability of the entire complex sentence, it poses a serious problem to current theory of selectional restrictions which appears to have no way of handling selectional restrictions across sentence boundaries.

Must is used in English to indicate a conclusion or a high degree of certainty. Must have is the past time analogue and can't its negation:

(72) There must be a hundred people here.

(73) There must have been a hundred people here.

(74) There can't be a hundred people here.

Unlike may, must is never negated. When not follows must, and even when it's merged with must in musn't, what is negated is the following infinitive, not must itself. A sentence like (75)

(75) John must not know the answer.

is consequently only two-ways ambiguous:

(76) It must be the case that John does not know the answer

(77) It is necessary that John does not know the answer.

The conclusive must is not used in interrogatives so that, for example, the tag for (78) is often (79).

(78) You must be out of your mind

(79) Aren't you?

Similarly, (80) is ill-formed.

(80) *Must you be out of your mind?

Will is of course treated in traditional grammar as above anything else the marker of future tense, along with shall. More commonly, will suggests willingness or agreement. It is formally distinguished from the future will in that in this sense will can occur in conditional clauses:

(81) If he'll come tomorrow, the matter will soon be settled.

Analogous to *may* and *must*, a similar ambiguity obtains for *will*, most commonly in negation.

(82) John will not confess his crime

is ambiguous:

(83) It will not happen that John will confess his crime

(84) John refuses (will not agree) to confess his crime.

Again, the ambiguity can also be sought in differences in the scope of the particle *not*. In (83) it is the infinitive *confess*, and in (84) the modal itself that is being denied.

Of the modals that express the idea of necessity or obligation (= moral necessity) in various degrees, we can recognize *must* and *ought to* (*should*).

Must expresses a degree of constraint that is felt as too strong to permit escape--necessity, in other words. In this sense, it may refer to the future; its analogous past time is *had to* and its negation *needn't* (or *don't have to*):

(85) I must go now.

(86) I had to go then.

(87) I needn't go now.

Analogous to permissive *may*, the conclusive *must* is not interrogated. The question corresponding to (88) would be something like (89).

(88) He must be an engineer

(89) Are you sure he is an engineer?

Also parallel to *may* is the fact that sentences containing *must* can be shown to be systematically ambiguous; those that are not readily apparent are in fact so under different circumstances.

Should and *ought*, no longer felt as inflected forms of *shall* and *owe*, are now used to express a degree of constraint

that is felt as escapable, as pointed out earlier. Their past time analogies are should have and ought to have; their negations are shouldn't and needn't respectively. Contrast should and the colloquial have to.

(90) I have to study tonight.

(90) implies that no escape from the task is in sight. Escape may later be found, but this is another matter. (91) implies that escape from the task is quite possible.

(91) I should study tonight.

(92) implies that no escape was found and (93) implies that escape was actually found.

(92) I had to study last night.

(93) I should have studied last night.

Ought has a much narrower range of meaning than should and it always leaves open the possibility of non-action, while must does not. We may thus attest:

(94) He ought to go, but he won't.

(95) *He must go, but he won't.

Like musn't, oughtn't is the negative form of ought only morphologically. Logical negation of both must (obligation) and ought is needn't. Compare (96)-(99).

(96) I must go, but John needn't.

(97) Must I go? No, you needn't.

(98) I ought to go, but John needn't.

(99) Ought I to go? No, you needn't.

Semantically, mustn't and oughtn't do not negate the obligation to act, but express a positive obligation not to act. We may thus contrast (100)-(102) and (103)-(105).

(100) You must go.

(101) You needn't go.

(102) You mustn't go.

(103) You ought to go.

(104) You needn't go.

(105) You ought not to go.

This distinction can't be made with all other modals; we can't with will, for instance, differentiate between denying the futurity of acting and stating the futurity of non-acting.

Be to expresses a kind of constraint that grows out of arrangement, stipulation and expectation of various kinds. It is often a polite substitute for the more direct have to or the more brutal must.

In summary, we can observe that in contrast with non-modal verbs in which past tense forms and those that are used in indirect discourse are not distinct, English modals have the complication that not all the past tense forms are used simply to refer to past time. The past time analogies vary for one simple modal in its various meanings and are certainly not always the past tense forms. A further complication is that not all modals have past tense forms, chiefly because they refer to future time or express logical necessity. In considering what the past time analogies are for each modal, we note three possibilities:

(a) past time reference is made with all of the modals in one of their senses with have, e.g.,

(106) He can have been at home yesterday. (possibility)
(107) He may have come last week. (possibility)
(108) They must have done it then. (certainty)
(109) You ought to have come with us yesterday. (desirability)

(b) for some of the senses, the past time analogue seems to be past tense plus have:

(110) He would have done that for you. (volition)
(111) He could have gone. (permission)
(112) I could have done that if you had asked. (willingness)

But these forms are generally referring to events that failed to happen.

(c) must has entirely different verbs as past time analogies:

(113) John must go (now).

(114) John had to go then.

(115) John must have gone.

3. This section will be devoted to the problem of how modals are to be introduced and represented in the deep structure and to a discussion of the syntactic properties of modals in general terms. My point of departure will be Ross's paper, "Auxiliaries as main verbs," where he first argues that all Aux's belong to the same major category as true verbs and are to be introduced into the deep structure the same way other verbs are and that there is no Aux constituent in the deep structure. I assume in the absence of counterarguments that Ross's arguments are conclusively established. In the following, although I will be basing my discussion largely on the evidence from the "neutral" modal may, it will be easy to extrapolate and extend my arguments, if valid, to be generally applicable to all modals. The arguments presented below for the subject-embedding may, for instance, apply to the conclusive must and those for the may which takes a sentential complement also seem to apply to the necessity must.

We begin by observing that permissive may, when used as a performative, may be a true verb which has a first person subject, as in (116),

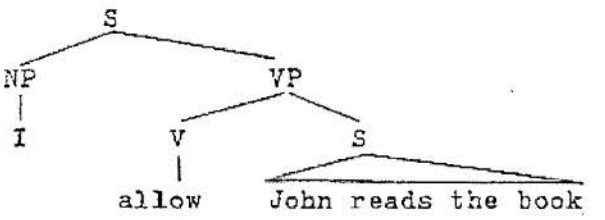
(116) John may read the book.

or unspecified subject as in (117).

(117) Cars may not park in this lot.

The deep structure of (116) would roughly be something like the following:[4]

(118)



---

[4] I leave open the question whether the deep structure in (118) is in fact the correct one or (116) is actually an instance of NP complementation. It may be noted, however, that (116) may well be an instance of VP complementation since the pseudo-cleft sentence is impossible:

*What I allow John is that John reads the book.

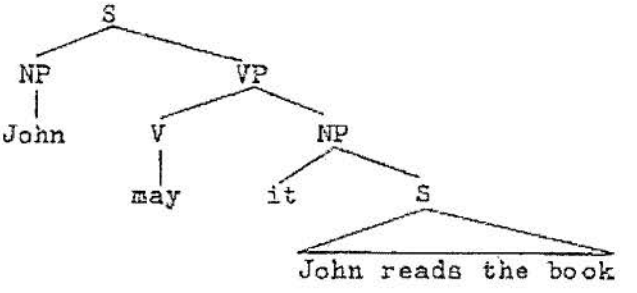
or

*What John may is to read the book.

Lexical substitution rule(s) will consult the DS, substitute may for the portion of tree dominating I allow (the details of which are not known to me); The Flip transformation then applies obligatorily to produce the correct surface form of (116).

An alternative, but much less plausible DS for (116) could be something like (119)

(119)



and Equi-NP deletion applies to the complement sentence to produce the correct surface form. A DS like this, being

bound up too much with the surface representation, seems to be wrong on two counts: (1) it assigns falsely a transitive reading to may; and (2) it fails to capture the fact that underlyingly may is a performative verb.

Note the switch of agents from active to passive sentences involving may (and modals in general): while in (120) the agent is the speaker, its active counterpart (121) implies that the agent is anything but the speaker.

(120) John may be examined by me.

(121) I may examine John.

Below are a class of simple may-sentences containing all three persons. On the right is exhibited the switch of persons involved. Matrix agent is the person giving the permission and constituent agent is equivalent to the embedding subject.

	Matrix Agent	Constituent Agent
(122) I may examine John.	He	I
(123) ?John may be examined by me.	I	I
(124) You may examine John.	I	you
(125) John may be examined by you.	I	you
(126) Mary may examine John.	I	Mary
(127) John may be examined by Mary.	I	Mary
(128) I may examine you.	He	I
(129) ?You may be examined by me.	I	I
(130) You may examine me.	I	you
(131) I may be examined by you.	I	you
(132) John may examine me.	I	John
(133) I may be examined by John.	I	John

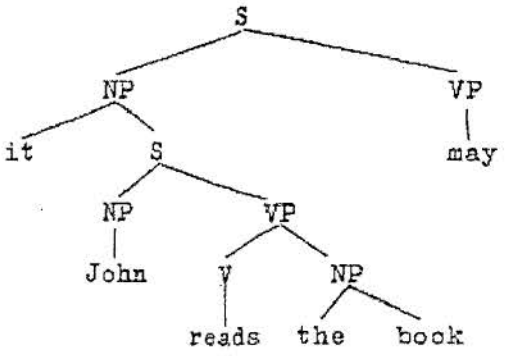
One thing that emerges clearly from the above comparison is that the second person you can never appear as matrix agent. Also the whole range of sentences containing all possible combinations of persons show that there is a constraint in English to the effect that no matrix agent can be identical to

a constituent agent. These sentences are marked with a question mark above.

May in the possibility sense might be an intransitive verb like seem, appear, happen, etc., so that the DS of (134) would be something like (135),

(134) John may read the book.

(135)



with the subject NP John being substituted for it by It-replacement and the rest of the embedded sentence being moved to the right and brought under the domination of the matrix VP yielding the correct surface form of (134).

It has been noted that the possibility may is never negated. A sentence like (136) is paraphrasable with (137) but not with (138).

(136) John may not read the book.

(137) It is possible that John will not read the book.

(138) It is not possible that John will read the book.

This characteristic of allowing only unidirectional negation is not to be found in verbs like seem, happen, appear, since (139) has the paraphrase (140) or (141).

(139) John seems not to be reading the book.

(140) It seems that John is not reading the book.

(141) It does not seem that John is reading the book.

There may be no explanation for the fact that only the embedded VP can be negated, given a DS like (135). Other seemingly unexplainable facts with regard to the possibility may are (142) and (143).

(142) it is never interrogated

(143) it is hardly passivized.

A DS like (135) in which it appears might be objected to on the ground that it is not a meaning-bearing element. Note, however, the presence of it may be available to account for sentences like (144) in a simpler way.

(144) It may be (the case) that John will read the book.

May, like seem, etc., does not permit the sentential subject to be moved to the front and topicalized, indicating that the extraposition transformation is obligatory for this class of verbs.

(145) *That John will read the book $\left\{\begin{array}{l}\text{may}\\\text{seem}\\\text{appear}\\\text{happen}\end{array}\right\}$.

If the matrix VP node is further expanded, we get the grammatical sentence (144) to (146).

(146) It may be strange that John will read the book.

Sentences (144) and (146) can be pseudo-clefted since the NP subject contains a sentence (147).

(147) What may be the case is that John will read the book.

and (148)

(148) What may be strange is that John will read the book.

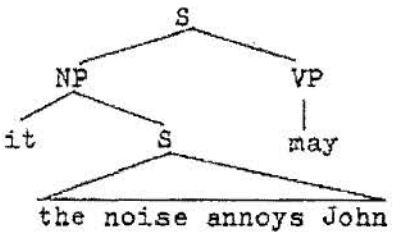
Consider now a pair of related sentences like (149) and (150).

(149) The noise may annoy John.

(150) John may be annoyed by the noise.

This pair of sentences are synonymous. If their DS is something like (151),

(151)



it can account for the synonymy of (152) and (153) (and hence (149) and (150)). Since the only difference between (149) and (150) is that the passive transformation has applied in the embedded sentence of (150) but not in (149).

(152) It may be that the noise will annoy John.

(153) It may be that John will be annoyed by the noise.

If, on therother hand, may is a verb like condescend, taking a sentential complement or a transitive verb like want taking a sentential object, we would expect (149) to exhibit some difference in meaning since the deep subject of (149) would be noise; that of (150) would be John.

This concludes my discussion of the syntax of the English modals. There are several problems I've not addressed myself to. I've not committed myself to an explanation, for instance, of why the permissive may is greatly weakened under the passive transformation, if it is possible at all. One may explain this by saying that the passive takes place in the complement sentence of the modal. In the case of possibility may, which has a DS like (118), the new subject is simply raised to give the passive surface form. The permissive may must, however, undergo lexical substitution rules and the Flip transformation. Since

the idea of permission is difficult to associate with the object of the embedded sentence, the passive sentence, thus produced, is strange in this reading.

Bibliography

Bochénski, I. M., 1951. Ancient Formal Logic, North Holland Publishing Company, Amsterdam.

Diver, William, 1964. "The modal system of the English verb," Word 20.322-52.

Joos, Martin, 1964. The English Verb, Madison: University of Wisconsin Press.

Long, Ralph B., 1961. The Sentence and its Parts, University of Chicago Press.

Ross, John R., 1967. "Auxiliaries as main verbs," unpublished paper.